A Socio-Cognitive and Computational Model for Decision Making and User Modelling in Social Phishing

Sunil Chaudhary[1], Eleni Berki[1, 2] Linfeng Li[3], Juri Valtanen[4], Marko Helenius[5]

[1]Faculty of Natural Sciences, University of Tampere,
Kanslerinrinne 1, Pinni B, 30014, Tampere, Finland,
chaudhary.sunil.x@student.uta.fi, eleni.berki@uta.fi

[2]Department of Computer Science and Information Systems,
University of Jyväskylä,
P.O. Box 35 (Agora), 40014, Jyväskylä, Finland

[3]Beijing Institute of Petrochemical Technology, Information Engineering College,
19 Qingyuan North Rd, Daxing, Beijing, China,
lilinfeng@bipt.edu.cn

[4]Faculty of Education, University of Tampere
Åkerlundinkatu 5, FI-33014 University of Tampere, Finland
valtanenjuripetri@gmail.com

[5]Department of Pervasive Computing, Tampere University of Technology,
P.O. Box 553, 33101, Tampere, Finland
marko.t.helenius@tut.fi

Abstract

Systems software quality, and system security in particular, is often compromised by phishing attacks. The latter were relatively easy to detect through phishing content filters, in the past. However, it has been increasingly difficult to stop more recent and sophisticated social phishing attacks. To protect the citizens from new types of phishing attacks, software quality engineers need to provide equally sophisticating preventive technology that models people's reactions. The authors considered the behaviour of people on the Internet from a socio-cognitive perspective and deduced who could be more prone to be spoofed by social phishing techniques. The authors herein propose a computational and interdisciplinary metamodelling methodology, which can assist in capturing and understanding people's interactive behaviour when they are online. Online behaviour can reveal Internet users' knowledge, information, and beliefs in a given social context; these could also constitute

significant factors for trust in social phishing circumstances which, in turn, can provide valuable insights and decision making meta-knowledge for recognition of potential victims of phishers. The proposed modelling approach is illustrated and explained using real-life phishing cases. This meta-model can i) help social computing and phishing researchers to understand users' trust decisions from a socio-cognitive perspective, and ii) open ways to integrate artificial intelligence design techniques within software quality management practices in order to protect citizens from being spoofed by social phishing attacks. Thus, this software design quality approach will increase system security as a proactive maintenance strategy.

Keywords: security, social phishing, trust, user behaviour modelling, metamodel, finite state machine (FSM), usability and learnability, social cognitive theory, software quality engineering.

1.0 Introduction

Social engineering is the psychological manipulation of people's vulnerabilities to produce a desired effect based on people's emotions and general predictive behaviour [1]. Social engineering techniques take advantage of people's cognitive biases and misconceptions or misinformation. In the security field, this is also referred as the art and science of human hacking [2]. Currently, social engineering is one of the most prominent methods used to conduct phishing attacks. The main reasons for this are: i) social engineering is relatively easy to apply compared to other hacking techniques and ii) there is no limit to social engineering techniques as long as there exists imagination and dark creativity to exploit different situations and contexts [3]. Such phishing attacks that employ social engineering tricks are also known as *social phishing* [4]. Social phishing is increasing and becoming more deceptive and sophisticated and, thus, difficult to be recognised at once [5]. For example, *socially-aware phishing*, *context-aware phishing*, *reverse social engineering* and *baiting* are types of intelligent and sophisticating phishing attacks.

Since the available anti-phishing technologies are not well-designed to protect against social phishing attempts, there is a need to improve the design and software quality of social computing technologies, so they, in turn, can help protect the citizens [6]. From the software engineering perspective, the quality of the software should be assured mainly by focusing on the users, i.e., considering usable security methods and strategies [6, 7]. Equally important is equipping the people with effective anti-phishing knowledge, skills and awareness [6, 8, 9]. Otherwise stated, by understanding the vulnerabilities of the potential (and ideal for social phishers) victims, the researchers and practitioners in the areas of social computing, usable security, security education, training and awareness should be able to integrate the existing adaptive techniques (e.g. [10, 11, 12]) to protect people from being spoofed by social phishing attacks.

This research study is led by the research questions: RQ1. '*Who are more prone to be spoofed by social phishing techniques?*' and RQ2. '*Could online users' behaviour and decision making be modelled?*' To answer these questions, the authors need to study the role of beliefs and contexts in the 'netizens' decision making process in the social phishing context, and upon this propose a logic model of the netizens' behaviour. This model can help social computing researchers understand people's trust decisions from a behavioural perspective, and eventually point towards adopting existing artificial intelligence techniques to adaptively prevent people from being spoofed by social phishing attacks.

In this paper, the authors first review the related user behaviour research in phishing. Second, the authors propose the way of defining a model of ICT users' behaviour towards social phishing attacks. Last, the authors carry out a test to verify the proposed model.

2.0 Literature Review and Related Work

There exist user-related studies on phishing attacks. Some researchers conducted usability research on end-users: Zhang et al. observed users when they were using different types of anti-phishing toolbars, and discerned their usability defects [13]. Li conducted usability evaluation studies and i) concluded on what information and knowledge anti-phishing toolbars should convey and ii) gave valuable advice on how to present this security-related information in a usable way [7].

Besides the usability research, some researchers delved into phishing problems from the users' behaviour perspective. Some researchers conducted research to find out about modelling of phishing in the social phishing context [7, 14, 15] and proceeded in constructing some models (or meta-models) of phishing under certain abstract conditions and circumstances. For example, Jakobsson conceptualised and implemented a graphical representation to capture and model the essence of phishing attacks. Such models can help in detecting the vulnerabilities of a system and determine suitable defence mechanisms for the users. [14]. Similarly, Li and other researchers built a mathematical model to depict users' behaviour in the phishing context [7]. In this model, it was emphasised that the appropriate available knowledge is the key factor to impact the decisions of online users and thus influence their choices and beliefs in the content of a phishing attempt. Finally, in the work of Dong et al., a model for visualising the interaction between user and phishing was designed [15]. This type of model can help security professionals in determining the mismatches between users' perceptions of phishing attacks and the attacks in reality, so that the latter can be captured and handled through the design of anti-phishing applications and suitable education.

Although the above mentioned studies contribute to the phishing research in a meaningful way, they significantly lack in addressing the problem of social phishing by considering substantial details. Further, in order to understand the end users' interactive behaviour in the socio-cognitive context of this particular type of

emerging phishing attacks, some user behaviour research needs to be further conducted. The user behaviour modelling can actually be proved helpful in viewing and defining a sequential order of the decision making process. Therefore, one alternative to model online citizens' behaviour is to utilise decision-based making theories, which could also be useful as supportive references for social phishing research. For example, a *decision tree* is one technique that is simple to understand since it describes the decision making process using a flow-chart-like model [16, 17], which consists of three types of nodes, decision nodes, chance nodes, and end nodes. A decision tree can sequentially depict how factors can affect the final decision. Adopting a more abstract computational perspective, a decision tree can be considered as a deterministic finite state automaton, in which the sequence of the factors is deterministic and each node of a decision tree is a state in the automaton. However, the rather fixed properties of the deterministic and non-deterministic finite state automata are not enough for so many different users' behaviour modelling. This is so mainly because for different individuals (with different beliefs, knowledge and value systems) the sequence of the factors is not and cannot always be the same [7] due to the richness of the socio-cognitive context and situations and the different trust requirements and dependencies.

In previous social and cognitive context studies [18, 19, 20, 21], trust has not been considered as only a mental attitude/attribute or a pure internal belief. Instead, the concept of trust is described as consisting of three basic elements: *a mental attitude, a decision to rely upon the other,* and *behaviour*. Moreover, a mental attitude represents a belief from the evaluation of the agent's trustworthiness, and a prediction based on the agent's willingness and ability to produce some effects [20]. A decision to rely on the others refers to the intention to delegate the production of a desired goal [20]. Behaviour means to take actions to trust another agent and build a practical, informational relation between the parties [20]. In this way, trust is described as a framework, whose elements are isolated from each other. For example, Durante presented that the internal attribution of trustors and the environmental attribution of trustees affect the trust [20]. Another example is provided by other researchers, who also tried to present the trust model as capital and studied the cognitive dynamics from the capital perspective and point of view [21]. Furthermore, Castelfranchi et al. [19] built a tree model to describe how users make trust decisions. In this tree model, the researchers defined the different weighed value(s) for different factors of trustees. The use of different weights can result in finding how much the different factors can affect the trustors' attitude on trust. Although these studies apply the trust theory from the social and cognitive perspectives, the models themselves are not able to define how (or how much/far) the different internal and external attributions affect each other and correspondingly affect the final trust decision. Considering the previous, the authors proceeded to a mathematical (computational) method on how to model the behaviour pattern of an online user.

3.0 Behaviour Modelling Methodology for Online Users

In order to define as many factors as possible considered in the social phishing research, the authors proceeded to an analysis of the phishing context from a *socio-cognitive perspective*. In the social phishing context, trustors are the potential victims who receive phishing information, and trustees are the phishers associated with their own phishing information. To build a trust relationship, trustees should present their *internal and external attributions* so that trustors may believe that trustees can be trusted. In a socio-cognitive theoretical framework, this means that these attributions are induced from the trustors' perspectives and affect the trustors' decisions [19, 20, 21, 22]. Although these studies highlighted the combined effect of competence and sincerity on the trustworthiness of information, the online security research field still lacks studies on how and on which order the attributions affect each other in their models.

The behaviour modelling methodology introduced in this paper is based on the *finite state automata theory* [23]. This theory represents a dynamic and computational modelling approach, which is able to describe a sequence, a selection, a multiple selection, and a repetition (or iteration) of events and/or attributions through transitions of states, which depict the situational context of phishing. The finite state automaton (FSM) has a limited, finite number of possible states. It has initial states, final states and current states. At each change of states, a deterministic or/and non-deterministic input is given, and the next state is correspondingly transitioned. The new state depends only on the current state and the symbol input. Regarding the representation of the FSM, the conventional notation is also followed, that is: a circle represents a state, an arrow represents state transition and the arrow label indicates the input value corresponding to the transition. The initial state is usually represented by an arrow with no origin pointing to the circle and the final state is drawn by a double circle. Next the authors illustrate the above through examples in which they define the relationships among these internal and external factors that outline and determine the characteristics of the potential human victims, who should rather be considered as a more-prone-to social phishing attacks group of citizens.

3.1 Modelling Internal Attributions

The internal attributions utilised and considered for modelling here include [20]:

- *Competence*: Trustee's qualities such as skills, expertise, and knowledge needed to perform the task.
- *Willingness*: Trustee's intention and readiness to perform the task.
- *Persistence*: Trustee's steadiness in the intention to perform the task.
- *Dependence*: Trustor's belief that it is either necessary or preferable to rely on the trustee in order to obtain a goal.
- *Fulfilment*: Trustor's belief that the goal will be achieved due to the trustee.
- *Motivation*: Reasons that persuade the trustee to adopt the goal.

The authors use a phishing email as an example and subsequently analyse the case utilising the above attributes:

Case 1: *Your flight is cancelled; please transfer100 euro to bank account xxxx-xxxx-xxxx-xxxx as collaterals to reserve the seat for your next flight. We will refund the money back to you after your trip.*

In case 1, this is how the trustee(s), i.e. the fraudster(s), presented or implied their internal attributions, factors in the social context of phishing. The *competence* is that the agency is competent on the ticket reservation; the *willingness* is that the agency is going to give this offer to every passenger, whose flight is also cancelled; the *persistence* from the message implies the money is required based on the flight company's regulation; the *dependence* refers to paying the money is one preferable way to reserve the flight; the *fulfilment* means only the flight agency knows how to reserve the flight ticket in this special occasion; and the *motivation* of this message is that this is a part of the agency's commitment.

To apply the finite state automata theory to model the above internal attributions, one state and its subsequent states associated with their corresponding inputs need to be defined. In the *social-cognitive theory*, *internal and external attributions* are the factors that impact the attitude of trustors. Herein, the authors firstly consider the internal attributions as the direct input of the model. The internal attribution of trust depends on the evaluation of the *trustee's qualities and defects*. An evaluation result of different internal attributions can be selected as inputs, and the different inputs can make the current state to the different subsequent states respectively. Next, the internal attributions are discussed situation wise.

Situation 1: Herein, the authors firstly consider the internal attributions as the direct input of the model. If a person receives a piece of phishing information and believes that s/he has adequate knowledge on phishing prevention, the person has no intention to rely on the phishing information (trustee). This description is able to be defined in finite state machine as Figure 1. In the figure, the S_0 is the state that a person receives a piece of phishing information, i_1 stands for the input that the person believes that s/he has adequate knowledge on phishing prevention. S_1 represents the state that s/he has no intention to rely on the phishing information.

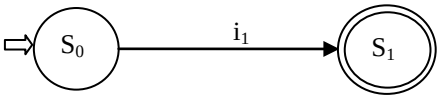

Figure 1: Transferring from one state to another state with internal attribution as an input

Situation 2: Many different internal attributions can be selected as inputs, and the different inputs can make the current state to the different subsequent states respectively. For example, in Figure 2, the S_0 is the state that a person receives a

piece of phishing information, i_1 stands for the input that the person believes that s/he has adequate knowledge on phishing prevention, and i_2 is that s/he has an intention to achieve the goal (e.g., update the security protection of their online banking services) in the phishing information. The state S_1 to be transferred with the input i_1 means s/he would not intend to rely on the phishing information, and the state S_2 to be transferred with the input i_2 indicates that s/he intends to rely on the phishing information.

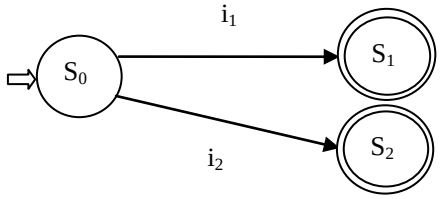

Figure 2: Transferring state from one state to two different states with two internal attributions as inputs

Situation 3: When defining the ways to transfer to the target state, there might be many possible cases. Firstly, it is possible to transfer from a source state with different inputs, e.g., in Figure 3 the state S_1 can be reached from S_0 with different inputs i_1 and i_2. For example, the S_0 is the state that a person receives a piece of phishing information, i_1 stands for the input that the person believes that s/he has adequate knowledge on phishing prevention, and i_2 is that s/he has no intention to achieve the goal (e.g., up-date the security protection of their online banking services) in the phishing information. The state S_1 to be transferred with both inputs means s/he would not intend to rely on the phishing information.

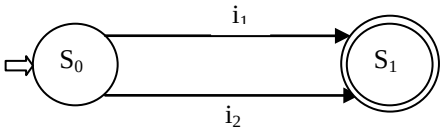

Figure 3: Transferring from one state to another one state with two internal attributions as inputs

Situation 4: It is also possible that a subsequent state can be transferred from different previous states. For example, in Figure 4, a person receives a piece of phishing information directly (S_0) or this person is asked for help from someone who claims to be his or her friends or relatives (S_0'). From the both source states, the person could believe in s/he has adequate knowledge (i_1) so that not to intend to rely on the phishing in-formation (S_1).

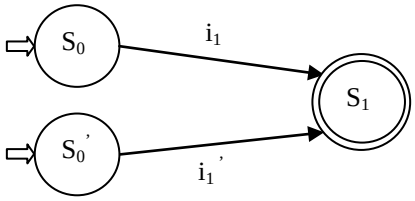

Figure 4: Transferring from two states to another one state with two internal attributions as inputs

In this study, the authors assume that all the parties in the phishing context share the same scale to measure the trustworthiness; also when a trustor evaluates the trustworthiness of a trustee, a trustor may use his/her own expectations as the benchmark value to compare with the corresponding value internal attributions of trustees. If a benchmark value is different from the value of the corresponding internal attributions, the trustor may have positive or negative attitude towards the corresponding internal attributions. The formula would be as follows:

$$w_k = A_k - E_k \text{ (k=1, 2…, 6).} \quad \textbf{(1)}$$

In this formula, A_k is the presented value of the k^{th} internal attribution, E_k is the expectation value of a trustor towards the k^{th} internal attribution, and w_k denotes the trustor's attitude towards the k^{th} attribution. If the presented value of an internal attribution is higher than a trustor's expectation value, the trustor will have a positive attitude on the corresponding internal attribution. The previous research on emotion analysis and deep learning *on the text* [10, 11, 12] can be applied to quantify A_k and E_k.

Previous research studies have observed that the adaptability of agents in the trust relationship [20] and the feedback effects of some internal attribution of trustees [19] are also emphasised while modelling the trust relationship. Therefore, when defining a user behaviour model in social phishing context, feedback (i.e., past experiences) should also be considered.

For example, in the aforementioned case 1, a person receives the phishing email (S_0), and s/he evaluates the phishing email (S_1) utilising the adequate knowledge s/he possesses (i_1). After his/her evaluation, s/he detects *trustee's defects*, which is true in actual. Thus, s/he makes a correct decision protecting himself/herself from falling for the attack, and the positive feedback from this experience is given as input (i_1) when the same state (S_1) is reached next time. This means that the same input (i_1) will be used for the same state (S_0) in the future. Otherwise, the negative feedback is given when the same state (S_1') is reached again, which means the same input (i_1') is not taken into account and other positive-feedback input, for example (i_2), is going to be used. In addition, the accumulated influence of the same feedback should be considered in the model, i.e. the feedback impacts the

attitudes towards the internal attributions in an accumulated way, and the formula (1) is now improved as follows:

$$w_k = A_k - E_k + w_k' \text{ (k = 1, 2…, 6).} \quad \textbf{(2)}$$

In the above formula, w_k' denotes the adjusted value according to the feedback from the past experiences, which means the w_k' can be calculated with the following formula:

$$w_k' = A_k - E_k' \text{ (k = 1, 2…, 6).} \quad \textbf{(3)}$$

In the above formula, A_k is the presented value of the k^{th} internal attribution, and E_k' is the adjusted expectation value of a trustor towards the k^{th} internal attribution, which is changed according to the feedback from the past experiences. Apparently, if the feedback is positive, E_k' is changed to a smaller value than the previous time. Otherwise, $E_{k'}$ is changed to a bigger value than the previous time.

3.2 Modelling External Attributions

External attributions refer to positive and/or negative environmental conditions, including opportunities, resources, interferences and adversities [18, 20]. If the same case 1 is considered, the environmental conditions can be, for example, that the phishing message is received during a festive season when people need to travel; or it is bad weather, such as a stormy or a foggy day when flights are often cancelled due to low visibility; or the phishing email recipient has a busy schedule with no time to check the credibility of the email; and the list can go on. When a person receives such a phishing message, s/he may refer to these environmental conditions as well. Similar to the internal attributions, the external ones are also considered as the inputs for user behaviour models. The only difference in the way the external attributions are treated is, the way to express how the external attributions affect the trust attitude.

For the internal ones, the trust attitude is based on trustors' expectation value of certain internal attributions. However, in their modelling method, the authors consider that the external attributions may respectively have an effect according to the perceptions of different individuals. For example, someone may consider online social tools as a trustworthy resource (one type of external environment), but others may not think of that due to the different perception and possible negative or positive experiences regarding the security and privacy of these online social tools. Therefore, when the authors define how the trust attitude is affected, they only use the perception to depict the trustors' attitude on external attributions. This means that when a person perceives an external attribution as a positive one, the corresponding attitude has a positive value. Otherwise, the corresponding attitude has a negative value. This could be written as follows:

$$W_k = P_{k\,(}\mathrm{k} = 1, 2\ldots, \mathrm{n}). \quad (\mathbf{4})$$

In formula 4, Wk is trustors' attitude on the k^{th} external attribution, and P_k is the trustors' perception on the k^{th} external attribution. Same as in internal attributions, the feedback effects also apply to external attributions. The trustors' attitude value regarding the feedback on the k^{th} external attribution is given as follows:

$$W_k = P_{k+}\, W_k{}'\,{}_{(}\mathrm{k} = 1, 2\ldots, \mathrm{n}). \quad (\mathbf{5})$$

In formula 5, W_k' is the feedback value of the k^{th} external attribution from the last time, which equals the adjusted perception (P_k') on the k^{th} external attribution as follows in the next section.

3.3 Modelling User Behaviour: To Trust or Not to Trust?

The authors model user behavioural patterns mainly focusing on how people consider the internal and the external attributions. Therefore, modelling of user behaviour proceeds to combine the possible states (behaviour steps) and inputs (internal and external attributions) to describe the whole process of how the users' trust decisions are made step by step. In this case, the resulting model of user behaviour is *different respectively*. This is because the different expectation value of internal attributions results in the different trust attitude on the internal attributions, and the different perception of external attributions leads to the trust *attitude respectively*. Instead of giving a specific user behaviour model, the authors only present a modelling methodology described as follows:

- Every input is given a certain weighted value so that to explicitly add a trust attitude towards an internal or an external attribution.
- The inputs associated with their weighted value are selected, defined and added between two states.
- The order of state transitions is defined, i.e., to define the initial state, the final state(s) and the possible reachable/reached states transferred between the initial state and the final state(s).

In this way, the model finally ends up with one of the final states where a trustor makes a decision whether to trust or not to trust, in order to proceed to the next action. With this modelling methodology, the resulted model is a directed and weighted graph. Based on the definition of the weighted value of each input, the authors define that the path with the biggest weighted value is the most vulnerable mental model to social phishing attacks.

4.0 Modelling of Social Phishing Cases

Again, let us consider a real life phishing case that happened in China and analysed in detail by the researchers of the article in reference [24].

Case 2: *On the 9th of February 2014, soon after the 2014 Chinese New Year, the police in Guangdong province launched a raid on saunas, karaoke bars and other venues of ill repute in Dongguan, a city in Guangdong province of China, famous for manufacturing and a highly developed sex industry. The police detained 67 people and shut down 12 venues. The news should have confirmed that the sex industry is not protected by the law in China, however. Some people got phished because of (the reporting of) these social events happening. According to the text from a newspaper published in Guangdong, some victims reported that they received phishing SMS messages saying: "Dad, I have been caught by the police when I played in Dongguan last night, please transfer *** yuan as bail to the bank account ******."*

In the above case 2, a person receives the message claiming to be from a close relative and asks for money to bail out (state: S_0). In this simple case, let us take a look at how the receiver will make the decision on trusting the message or not.

Firstly, the authors assume that the content of the message is considered and checked against the recipient's knowledge. This means that the relationship should be guaranteed from the attributes of the message (input: ii), e.g., the subscription number of the sender's phone, the sender's accent, and the certain implicit behavioural patterns. If these internal attributions in the message present higher value than the receiver's expectation, receivers may give positive trust attitude towards these attributions (weighted attitude: w_{k1}). The next state (state: S1) in the mind of the receiver is to consider the external attributions (input: i_2), e.g., the various sources of news related to the content of the message, his/her children or relatives are actually visiting the said city, their children or relatives have habits of gambling and visiting similar venues, their children or relatives are out of the mobile phone range, and so on. If the receiver believes that the message content as described should have happened in the current occasion, the receiver's attitude would be towards trusting the message (weighted attitude: W_{k2}). After the evaluation of internal and external attributions, the receiver makes a decision (state: S_2).

It is also possible that after the message is received, the receiver firstly considers the external environment, which is just right at the time after the sex raid in Dongguan (input: i_1'). Then the receiver has a certain attitude on the external attributions (W_{k1}'), the environment (state: S_1'). Compared to the state S_1, the receiver's mind may be affected by the external attributions, the state S_1' is, therefore, different from his/her mental perspective. Regarding the receiver's next move, it can be assumed that s/he may believe that s/he knows the sender so well that it is assured that the sender is caught in the sex raid. Therefore, when the receiver considers the internal attributions of the message (input: i_2'), s/he has a lower expectation value on these internal attributions, and give higher weighted value on trust attitude (w_{k2}'). After the evaluation of the internal attributions, the receiver may make another decision (state: S_2). Both the behavioural patterns are depicted in Figure 5. From the figure, one can easily compare and find out the most

vulnerable behavioural pattern, i.e. the path in the diagram with the biggest weighted value of trust attitude.

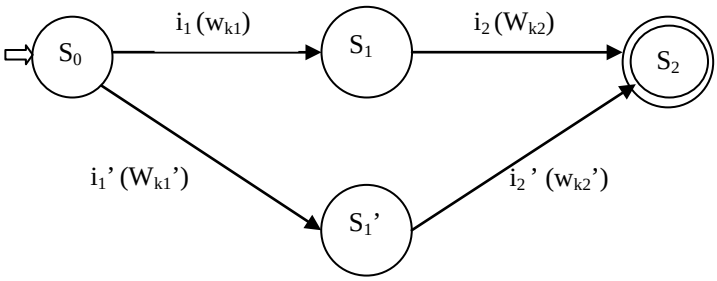

Figure 5: One example to show the user behaviour modelling in the social phishing context

In Figure 5, let us assume that the sum value of trust weight is $\sum W$, now

- If the trustor (i.e., the SMS receiver) chooses the first path, then $\sum W = w_{k1} + W_{k2}$ (where $k_1=1...6$; $k_2=1...n$)
- If the trustor chooses the second path, then $\sum W = W_{k1}' + w_{k2}'$ (where $k_1'=1...n$; $k_2'=1...6$)

In order to trust the mobile message, the sum value of trust has to be positive, i.e., $\sum W>0$. *The higher will be the sum value of trust, so will be the possibility that the trustor will respond to the phishing attack.* For the zero and negative values, i.e., $\sum W\leq 0$, the chance that the trustor will respond to the SMS will be minimal.

Humans have always been the weakest link in information security and, thus, the main target of deceitful attacks. If the question is why humans are easily deceived one can find many personal factors and reasons. Among the first that come to mind are: Socio-cultural issues such as a person might wonder how s/he will be remembered (was s/he helpful or not) or a person might not want to insult the other person (follows a high moral code), or a person thinks this is a good relationship investment (since social relationships are valued high), or a person is in a much lower social/power status (and thus does not dare to reject requests from others).

It constitutes a great challenge to consider assumptions of perceptual thresholds of multi-state positions (see e.g. [25]) to essentially argue why people lie and deceit [26], and why people trust and how they make trust decisions. Fareri et al., arguing on computational substrates of social value in interpersonal collaboration, support that ‘our brains reward us for taking the risk to trust’ [27]. Others support that people feel guilty if they do not trust other people relationships, businesses, governments etc. [28], and that this trust to strangers, even when it does not make

sense, seems to be the main reason that those bank scams on the Internet continue to flourish [28]. This study supports the need for adopting theories of vulnerable behaviour detection in assisting social phishing victims. This represents a radical innovation in design thinking about the way information is processed and informed decisions are made. There might also be limitations in utilising these theories. Special cases could be people who have lost the ability to make own decisions, such as ill people. They may have special security needs. (see e.g. [29]).

Many adaptive techniques [10, 11, 12] of learning users' behaviour types and emotions can complementarily be applied to collect people's emotional states and consequent attitudes, so that someone could calculate each input value to find out the most vulnerable behaviour towards social phishing attacks. With this computational and dynamic approach, the resulted (meta)model is a directed and weighted graph, that can be verifiable and testable due to its formal semantics and syntax [30]. Following this specification model someone can finally reach one of the final states, where a trustor, in order to proceed, makes a decision whether to trust or not to trust. For instance, based on the definition of the weighted value of each input, the authors define and verify that the path with the biggest weighted value is the most vulnerable mental model to social phishing attacks.

This is the first time that weighted FSM modelling is used in the context of social phishing and security. Similar weighted FSM models have been applied in protocol specifications and performance analysis [31] and in speech recognition [32].

5.0 Summary, Conclusions and Future Work

As used in psychology, education, and communication, socio-cognitive models depict an individual's knowledge use and acquisition and show how it can be directly related to observing people within the context of their social interactions, experiences, and external influences. Based on the latter, the authors introduced a computational modelling methodology to describe how people's beliefs, knowledge and social context affect their trust decisions in the case of social phishing attacks. This methodology utilises knowledge from interdisciplinary areas, including analysis of online users' needs, theories about trust and trustworthiness and classic computational theories through deterministic and non-deterministic modelling. Through this conceptual computational modelling, researchers and practitioners should be able to investigate vulnerable behavioural patterns of social phishing.

This particular modelling can also be helpful to anti-phishing software designers because it can assist in learning the vulnerable human behaviour patterns and warn users when a spoofing scam exploiting users' behaviour vulnerabilities is detected. It is convenient to apply the behaviour model in software when implementing the social-context phishing prevention tools. This type of modelling that targets to assist in the learning of vulnerable human behaviour patterns could increase anti-phishing software tools' learnability through computational intelligence

techniques, such as machine learning. For instance, the software design of anti-phishing toolbars and related technology could outperform when implementing the related design knowledge. This work captures a new type of design thinking, rich and abstract enough to model critical user needs and details. This user behaviour modelling method will further be implemented and integrated with adaptive algorithms to support adequate technology for public awareness.

6.0 References

1 Harley D (1998). Re-floating the Titanic: Dealing with social engineering attacks, European Expert Group for IT Security.

2 Hadnagy C, Social engineering, The art of human hacking. Wiley Publishing Inc. 2011, USA.

3 Simms C (2016). Is social engineering the easy way in? BCS the Chartered Institute for ITNOW, 58 (2), 24-25

4 Jagatic T, Johnson N, Jakobsson M, Menczer F (2007). Social phishing, Communication of the ACM 50(10) 94-100.

5 Samani R & McFarland C (2014). Hacking the human operating system: The role of social engineering within cybersecurity. Technical Report, Intel Security McAfee.

6 Chaudhary S (2016). The use of usable security and security education to fight phishing attacks. PhD Thesis, University of Tampere.

7 Li L (2013). A contingency framework to assure the user-centered quality and to support the design of anti-phishing software. PhD Thesis, Uni of Tampere.

8 Sheng S, Magnien B, Kumaraguru P, Acquisti A, Cranor LF, Hong J, Nunge E (2007). Anti-phishing Phil: The design and evaluation of a game that teachers people not to fall for phish, proceedings of the 3rd Symposium on Usable and Security, Pittsburgh, PA, USA, 18-20 July 2007, pp. 88-99.

9 Kirlappos I & Sasse MA (2012). Security education against phishing: A modest proposal for a major re-think. IEEE Security & Privacy, 10(2), 24-32.

10 Bousmalis K., Zafeiriou S, Morency LP, Pantic M (2013). Infinite hidden conditional random fields for human behavior analysis, IEEE Transactions on Neural Networks and Learning Systems, 24, 170-177.

11 Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S (2001). Emotion recognition in human-computer interaction, IEEE Signal Processing Magazine 18(1) 32-80.

12 Mikolov T, Kombrink S, Burget L, Cernocky J, Khudanpur S (2011). Extensions of recurrent neural network language model, proceedings of International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, 22-27 May 2011, pp. 5528–5531.

13 Zhang Y, Egelman S, Cranor LF, Hong J, Phinding phish: Evaluating anti-phishing tools, proceedings of the 14th Annual Network & Distributed System Security Symposium, San Diego, CA, USA, 28 February - 2 March 2007.

14 Jakobsson M (2005). Modeling and preventing phishing stacks. Financial Cryptography and Data Security, Springer Berlin Heidelberg 2005.

15 Dong X, Clark JA, Jacob J, Modeling user-phishing interaction, proceedings of the Human System Interaction, Krakow, Poland, 25-27.5.2008, pp 627–632.

16 Zhang Y, Hong J, Cranor L, CANTINA: A content-based approach to detecting phishing websites, proceedings of the 16th International World wide Web Conference, Banff, Alberta, Canada, 8-12 May 2007, pp. 639-648.
17 Ludl C, McAllister S, Kirda E, Kruegel C, On the effectiveness of techniques to detect phishing sites, proc. of the 4th Intern Conf Detection of Intrusions & Malware, and Vulnerability Assessment, Lucerne, 12-13 July 2007, pp. 20-39.
18 Falcone R & Castelfranchi C, Social trust: A cognitive approach. Trust and Deception in Virtual Societies. Springer Netherlands 2001, 55-90.
19 Castelfranchi C, Falcone R, Pezzulo G, Integrating trustfulness and decision using fuzzy cognitive maps. LNCS 2692, Springer 2003, 195 – 210.
20 Durante M (2010). What is the model of trust for multi-agent systems? Whether or not e-trust applies to autonomous agents, Knowledge, Technology & Policy 23(3), 347-366.
21 Falcone R & Castelfranchi C (2011). Trust and relational capital. Journal of Computational and Mathematical Organization Theory 11, 402-418.
22 Paglieri F, Castelfranchi C, Pereira C, Falcone R, Tettamanzi A, Villata S (2014). Trusting the message and the messenger: feedback dynamics from information quality to source evaluation. Computational and Mathematical Organization Theory 20 (2), 176 – 194.
23 Sipser M, Introduction to the theory of computation. 3rd Ed., Cengage Learning 2012, pp. 31-41.
24 News.163.com (2014). The police raid in dongguan becomes a spoofing resource (Translated from Chinese), Retrieved on 5 July 2016 from: http://news.163.com/14/0213/03/9KUCEQKI00014Q4P.html
25 Dember WN & Warm JS, Psychology of perception. Holt Rinehart and Winston 1979, USA.
26 Ford CV, Lies! Lies! Lies! The psychology of deceit. American Psychiatric Press Inc. 1999, Washington, D.C.
27 Fareri D, Chang L, Delgado M (2015). Computational substrates of social value in inter-personal collaboration. The Journal of Neuroscience, 35(21).
28 Park, A (2014). We trust strangers, even when it doesn't make sense to do so. Retrieved on 5 July 2016 from Time Magazine: (http://time.com/103396/we-trust-strangers-even-when-it-doesnt-make-sense-to-do-so/).
29 Silberfeld M & Fish A, When the mind fails. a guide to dealing with incompetency. University of Toronto Press, 1994.
30 Berki E (2001). Establishing a scientific discipline for capturing the entropy of systems process models: CDM-FILTERS - A Computational and Dynamic Metamodel as a Flexible and Integrated Language for the Testing, Expression and Re-engineering of Systems. PhD Thesis, Nov 2001. Faculty of Science, Computing & Engineering, University of North London, London, UK.
31 Ionescu C, Berki E, Nummenmaa J. (2009). Applying Weighted Finite State Machines to Protocol Performance Analysis. D. Dranidis & I. Stamatopoulou (Eds). 4th SE Eur Workshop on (i) Formal Methods on Web Services (ii) Formal Methods for Agent-based Systems: IEEE Comp. Society, pp. 40-45.
32 M. Mohri and F.C.N. Pereira. (1998). Dynamic Compilation of Weighted Context-Free Grammars. In 36th Annual Meeting of the ACL and 17th International Conference on Computational Linguistics, vol 2, pp 891–897.